



## 识别Z玻色子喷注的卷积神经网络方法

李靖 孙昊

### Tag Z boson jets via convolutional neural networks

Li Jing Sun Hao

引用信息 Citation: *Acta Physica Sinica*, 70, 061301 (2021) DOI: 10.7498/aps.70.20201557

在线阅读 View online: <https://doi.org/10.7498/aps.70.20201557>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

利用神经网络识别高分子链在表面的吸附相变

Recognition of adsorption phase transition of polymer on surface by neural network

物理学报. 2019, 68(20): 200701 <https://doi.org/10.7498/aps.68.20190643>

基于深度卷积神经网络的大气湍流相位提取

Extracting atmospheric turbulence phase using deep convolutional neural network

物理学报. 2020, 69(1): 014209 <https://doi.org/10.7498/aps.69.20190982>

基于深度残差网络的高精度自然转换模拟方法

High-order natural transition simulation method based on deep residual network

物理学报. 2020, 69(20): 204701 <https://doi.org/10.7498/aps.69.20200563>

基于遗传算法优化卷积长短记忆混合神经网络模型的光伏发电功率预测

A hybrid model for photovoltaic power prediction of both convolutional and long short-term memory neural networks optimized by genetic algorithm

物理学报. 2020, 69(10): 100701 <https://doi.org/10.7498/aps.69.20191935>

变频正弦混沌神经网络及其应用

Frequency conversion sinusoidal chaotic neural network and its application

物理学报. 2017, 66(9): 090502 <https://doi.org/10.7498/aps.66.090502>

多频正弦混沌细胞神经网络及其复杂动力学特性

Multi-frequency sinusoidal chaotic neural network and its complex dynamics

物理学报. 2020, 69(24): 240501 <https://doi.org/10.7498/aps.69.20200725>

## 识别 Z 玻色子喷注的卷积神经网络方法\*

李靖 孙昊†

(大连理工大学物理学院, 大连 116024)

(2020 年 9 月 18 日收到; 2020 年 11 月 13 日收到修改稿)

高能物理中喷注识别任务是从背景中识别出感兴趣的特定信号, 这些信号对于在大型强子对撞机上发现新的粒子, 或者新的过程都有着非常重要的意义. 量能器中产生的能量沉积可以看做是对喷注的一种拍照, 分析这样产生的数据在机器学习领域中属于一个典型的视觉识别任务. 基于喷注图片, 本文探索了利用卷积神经网络 (convolutional neural networks, CNNs) 识别量子色动力学背景下的 Z 玻色子喷注, 并与传统的增强决策树 (boosted decision trees, BDTs) 方法进行了对比. 在本文利用的输入前提下, 三种相关的性能参数表明, CNN 比 BDT 带来了约 1.5 倍的效果提升. 除此之外, 通过最优与最差的喷注图与混淆矩阵, 说明了 CNN 通过训练学习到的内容与整体识别能力.

**关键词:** Z 玻色子衰变, 夸克, 胶子, 神经网络

**PACS:** 13.38.Dg, 14.65.-q, 14.70.Dj, 87.85.dq

**DOI:** 10.7498/aps.70.20201557

## 1 引言

在大型强子对撞机 (LHC) 上, 粒子在非常高的能量下进行碰撞, 研究人员可以通过各种探测器观测到末态粒子束 (喷注). 因为能量如此之高, 喷注在量能器上产生的能量沉积有可能发生重叠, 这样会形成一个胖喷注. 胖喷注有着丰富的亚结构, 可以用来研究产生它的相应的过程. 以 Z 玻色子衰变到双喷注为例, 胖喷注含有潜在的多个亚喷注结构, 相比之下, 普通的量子色动力学 (quantum chromodynamics, QCD) 过程产生的胖喷注则没有这样的结构. 从庞大的 QCD 喷注背景下识别出特定过程产生的喷注对之后进行喷注研究有着重要的意义, 这就是喷注识别. 大量的工作通过理解胖喷注的亚结构来提出识别的方法, 这些工作的综述可以查看文献 [1–3]. 在物理学中, 传统的方法是设计一些观测量, 然后对这些观测量的分布人为施加截断, 全部的截断构成了一个分辨器; 也可以

单独或者组合利用这些观测量, 通过增强决策树 (boosted decision trees, BDTs) 这种机器学习算法来进行分辨. 尽管这些观测量是人为精心构造出来的, 但是它们并不一定能充分利用到喷注所含的潜在信息.

近年来, 大量的工作尝试将机器学习方法应用于物理中的不同任务 [4–6]. 其中对于喷注识别的任务, 不同的机器学习模型有着不同的输入, 于是人们构造了喷注的不同的表示方法, 例如: 图片 [7–18]、序列 [19–22]、图结构 [23]、集合 [24] 等. 通过利用模型自动提取特征的功能, 更高维度、更复杂的信息被产生用来识别喷注的来源. 本文的识别任务是从 QCD 背景下识别出由高能 Z 玻色子衰变而来的喷注. 区别于人为构造的观测量, 我们直接利用了低维的四动量数据, 来探索这一原始数据所包含的有用的潜在信息. 喷注被看作是一张图片作为输入, 然后利用卷积神经网络 (convolutional neural networks, CNNs), 通过层层相连的卷积层, 来提取出不同维度的特征图, 依次作为输入传入下

\* 国家自然科学基金 (批准号: 11675033, 12075043) 资助的课题.

† 通信作者. E-mail: haosun@dlut.edu.cn

一层,最后通过全连接层输出信号与背景的概率,通过相对概率的大小,喷注图片被识别成信号或者背景.在不同深度的 CNN 进行训练比较之后,找出了对于这一过程,最精简与最高效的 CNN 结构.为了对比它们与传统方法的差异,还使用了 BDT 来进行识别.结果显示 CNN 模型的效果远超 BDT 的效果,也说明了在未来的标注识别研究中,这种结构及其变体蕴含的巨大潜力.

## 2 数据产生

选取 Z 玻色子衰变产生双喷注作为研究的信号,选取普通并且非常庞大的 QCD 喷注作为背景.利用 Pythia<sup>[25]</sup> 模拟了这两种过程,产生了部分子水平的数据,它们的截面分别为  $2.485 \times 10^{-9}$  mb 和  $2.507 \times 10^{-6}$  mb. 中心能量  $\sqrt{s}$  设置为 14 TeV, 横向动量  $p_T$  范围设置为 400—450 GeV, 赝快度范围<sup>[7]</sup> 设置为  $|\eta| < 1.6$ , 方位角的范围是  $|\phi| < \pi/2$ . 得到了末态粒子的四动量后, FastJet<sup>[26]</sup> 用来聚集喷注. 只保留横向动量大于 5 MeV 的末态粒子, 并使用 Anti- $k_T$  算法<sup>[27]</sup> 来聚集  $\Delta R < 1$  的粒子来产生胖喷注.  $\Delta R = 0.3$  的  $k_T$  算法被用来做再聚合产生亚喷注. 为了降低潜在事件的影响, 舍弃了横向动量不足胖喷注横向动量 5% 的亚喷注, 这也叫做喷注修剪<sup>[28]</sup>.

经过筛选后每个事件剩下的末态粒子, 也就是上面保留下来的组成亚喷注的粒子, 被用来产生喷注图片. 粒子的横向动量  $p_T$  作为权重来投影到由赝快度  $\eta$  和方位角  $\phi$  组成的二维平面上. 为了模拟真实的量能器, 将数据的精确度设置为  $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ , 单位小格构成喷注图片的一个像素. 位

于同一像素的粒子, 它们的横向动量之和组成该小格的像素强度. 最后产生的喷注图片尺寸大小为  $32 \times 32$ . 为了让模型最快地学习识别信号与背景, 对喷注图片做了预处理. 预处理可以大大减少训练的时间, 并且提升模型的识别效果. 图片的产生和预处理步骤有: 平移、像素化、旋转. 首先将喷注数据中的首要亚喷注 (有着最大的横向动量) 移动到  $\eta$ - $\phi$  平面的原点, 即喷注图片的中心, 然后再将喷注所含有的全部粒子投影到这个平面上, 最后将次要的亚喷注旋转到中心的正下方. 图 1 展示了 Z 玻色子与 QCD 的平均喷注图片, 次要的亚喷注清楚地显示在信号喷注图 1(a) 中, 而背景喷注在图 1(b) 中, 相应区域的像素强度更加平均与发散, 没有形成明显的次要喷注区域.

在分别模拟了一百万的信号与背景事件之后, 最后得到的信号与背景喷注图片的数量总共约为五十四万, 各占其中的一半. 在将它们顺序打乱后, 其中的 30% 作为测试集, 剩下的作为训练集. 验证方法采用了三折交叉验证, 每次将训练集分成 3 份, 选择其中的一份作为验证集, 其余的两份作为训练集.

## 3 研究方法

### 3.1 卷积神经网络 (CNN)

本文卷积块 (ConvBlock) 由一个卷积层, 一个批归一化层, 一个最大池化层组成. 为了保持输入的尺寸大小不变, 卷积层的填充数设置为 1, 卷积步长设置为 3. 在这样的设计下, CNN 可以有更深的结构. 为了防止模型过于复杂带来的过拟合, 在卷积块的最后添加了丢弃层, 有 50% 的概率丢弃

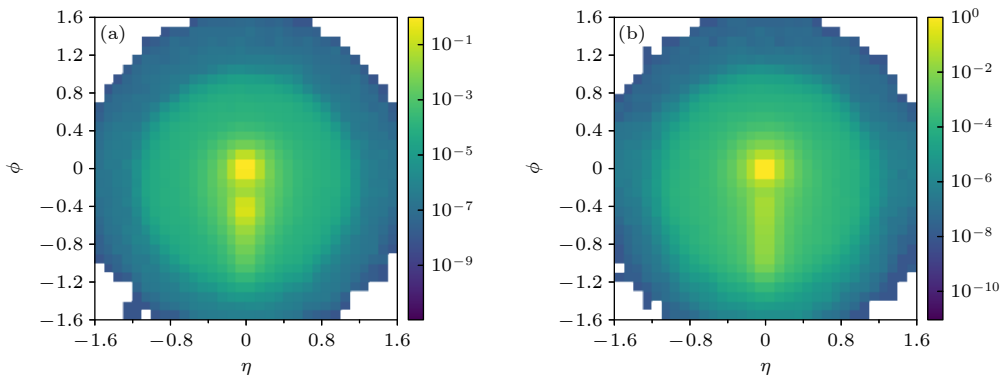


图 1 (a) 信号平均喷注图; (b) 背景平均喷注图; 横坐标  $\eta$  代表赝快度, 纵坐标代表方位角  $\phi$ .

Fig. 1. (a) Signal average jet image; (b) background average jet image.  $\eta$  and  $\phi$  represent pseudo-rapidity and azimuth respectively.

与之相连的特征图. 总共探索了四种 CNN 结构: 所包含卷积块的个数分别为 2, 3, 4, 5, 分别命名为 CNN 1, CNN 2, CNN 3, CNN 4. 展示了包含 4 个卷积块的 CNN 3 结构, 如图 2 所示. 所有的结构都是卷积块层层堆叠组成的, 最后加上一个全连接的分辨层得到输出. 随着层数变深, 中间得到的特征图通道数逐渐增多, 尺寸变小, 直到最后的单像素图. 除了这一种结构, 文献 [9, 17] 还探索了不同的结构. 在训练过程中采用了 Adam 优化算法, 学习率设置为 0.001, 同样为了防止过拟合采用了早停法, 在 20 个周期内如果验证集上的损失没有

下降的话, 训练将会终止. 此外使用了交叉熵损失函数. 模型由 Pytorch 搭建而成, 训练使用了 Pytorch 的高级封装 Skorch.

### 3.2 增强决策树 (BDTs)

为了衡量 CNN 的分辨效果, 将增强决策树作为基线, 聚集产生的胖喷注的质量、横向动量, 首要和次要亚喷注之间的  $\Delta R$ , 以及喷注形状 N-subjettiness 中的  $\tau_{21}$  作为它的输入, 图 3(a)—(d) 分别显示了它们的分布. 我们采用的是 Sklearn 中的梯度增强决策树 (gradient boosted decision tree,

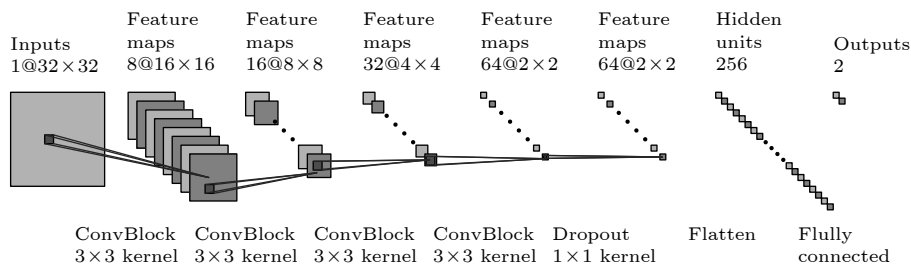


图 2 CNN 3 结构示意图, 产生这张图片的程序来自 [https://github.com/gwding/draw\\_convnet](https://github.com/gwding/draw_convnet)

Fig. 2. Architecture of the CNN 3. This figure was generated by adapting the code from [https://github.com/gwding/draw\\_convnet](https://github.com/gwding/draw_convnet).

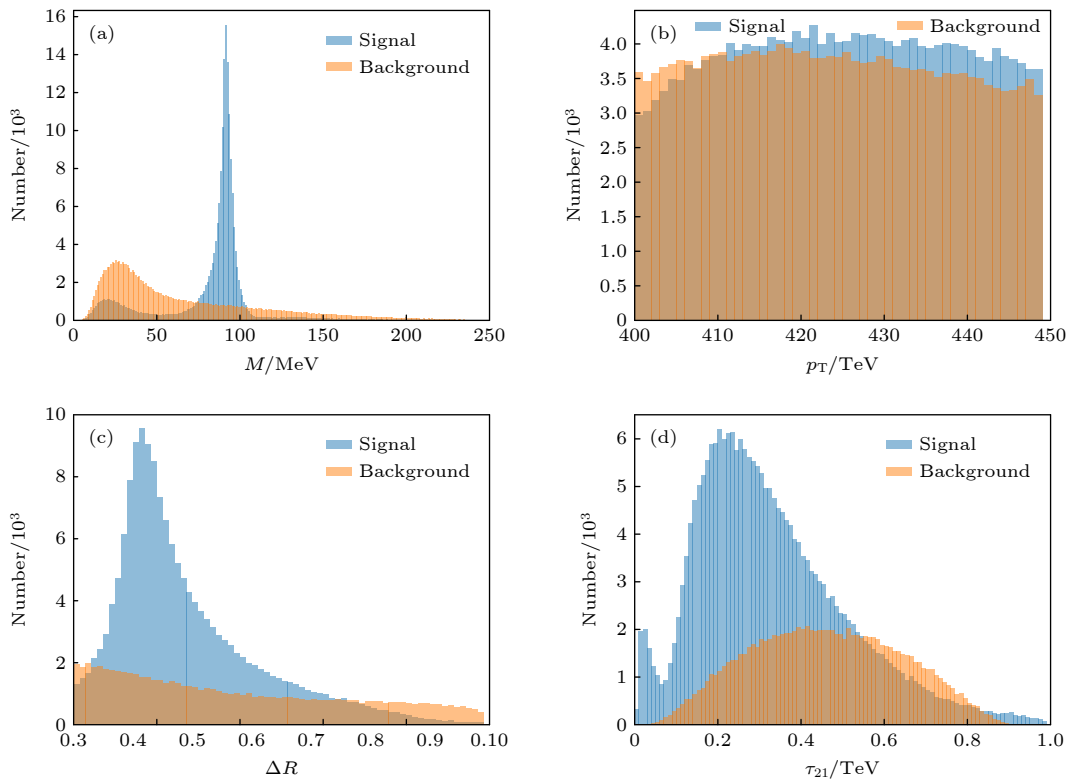


图 3 (a) 胖喷注的质量分布; (b) 胖喷注的横向动量分布; (c) 胖喷注含有的首要与次要喷注的距离分布; (d) N-subjettiness  $\tau_{21}$  的分布

Fig. 3. (a) Mass distribution of fat jets; (b) transverse momentum distribution of fat jets; (c) distribution of distance between leading and subleading subjects; (d) distribution of N-subjettiness  $\tau_{21}$ .

GBDT). 其中学习率设置为 0.1, 用来训练不同树的样本比例设置为 0.9, 每个树的最大深度设置为 3. 对于树的个数, 分别采用了 100, 200, 300 来试图找到最佳的设置. 需要注意的是, 这里出现的并不是全部的设置, 其他的设置可能会出现更好的模型, 这个将在未来进行探索.

#### 4 结果与讨论

在对实验结果进行分析之前, 先来说明一下本文用到的分析方法. 我们把识别得到的信号占真实信号的比例称作信号效率, 记为  $\epsilon_s$ , 错误地识别成信号的背景占真实背景的比例称作错误标记率, 记为  $\epsilon_b$ . 通常情况下当信号输出的相对概率大于 50% 的阈值时, 模型判断输入的喷注图来自信号. 可以利用接受者操作特性 (the receiver operating characteristic, ROC) 曲线来展示随着阈值的变化, 模型判断信号与背景的变化. 这里横坐标为  $\epsilon_s$ ; 纵坐标为  $1/\epsilon_b$ , 它也被称为背景拒绝效率. 一般来说希望模型在相同的信号效率下有更高的背景拒绝效率, 所以在 ROC 曲线图中, 曲线位于更高位置的模型, 认为它的表现更好. 除了 ROC 曲线, 还使用了其他三种参数: 曲线下的面积 (area under the curve, AUC) 来反映不同模型的性能差异, 这里的曲线指的是分别以  $\epsilon_b$  与  $\epsilon_s$  为横纵坐标作出的曲线, 与我们利用的 ROC 曲线不同; 准确度 (accuracy, ACC), 模型识别的正确率; R50, 在信号效率为 50% 的情况下背景拒绝效率的值.

接下来对得到的结果进行分析. 图 4 展示了前面提到的所有模型的 ROC 曲线, 表 1 展示相对应参数的值. 图中的 ROC 曲线表明了通过加深模型, CNN2, CNN 3, CNN 4 获得了比 CNN 1 更强的分辨能力, 但是 CNN 3 和 CNN 4 的分辨能力几乎是一模一样的. 这也可以反映出尽管 CNN 4 模型参数几乎是 CNN 3 的四倍, 但是无法再通过简单地加深模型来提升它的表现, 所以 CNN 3 是在这种将卷积块层层连接的设计下最精简的模型. 对于 BDT, 尽管 3 个表现参数随着树的个数增多而增加, 但是它们的对于分辨能力的影响极其有限 (甚至将树的个数设置为个位数, 但是它的表现依旧没有什么大的变化). 这说明对于 BDT, 此时限制模型分辨能力的因素已经不再是模型的结构, 而可能与我们选择的输入有关, 之后的工作可以尝试

更多种类的观测量加以验证. 需要注意并不是种类越多越好, 含有模糊信息的输入反而可能会误导模型使它的分辨能力下降. 对于我们选取的用来衡量模型表现的参数, 准确度的差异非常小, CNN 3 相较于表现最差的 CNN 1 与 BDT 1, R50 分别提升了大约 1 倍与 1.5 倍, 准确度达到了 0.8324.

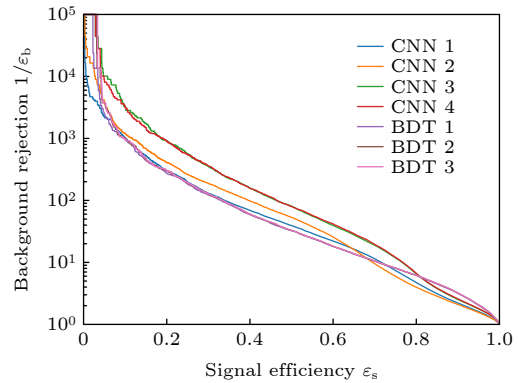


图 4 不同模型的 ROC 曲线

Fig. 4. ROC curves of different models.

表 1 用来衡量不同模型表现的性能参数

Table 1. Metrics to evaluate performance of different models.

模型	AUC	ACC	R50
CNN 1	0.8754	0.8150	39.1103
CNN 2	0.8688	0.8252	53.3583
CNN 3	0.8980	0.8324	80.6715
CNN 4	0.8993	0.8328	79.9350
BDT 1	0.8955	0.8337	32.5351
BDT 2	0.8963	0.8342	32.8072
BDT 3	0.8969	0.8346	33.0144

为了进一步了解 CNN 3 的分辨能力, 利用 softmax 函数将模型输出转化为对应类别的概率, 公式为

$$p(i) = \frac{e^{o_i}}{\sum_{j=1}^2 e^{o_j}}, \quad i = 0, 1, \quad (1)$$

其中,  $i$  表示输出神经元所代表的输入类别, 0 代表背景, 1 代表信号;  $o$  代表了神经元的本身的输出. 选取信号神经元来查看由不同类别的输入得到的输出分布, 如图 5 所示. 图中信号的输出大部分集中于 1 附近, 背景集中于 0 到 0.3 附近, 模型可以很好地将它们区分开来.

为了探索神经网络学习到了喷注所包含的哪些结构, 分别作出了最优与最差的信号与背景的喷



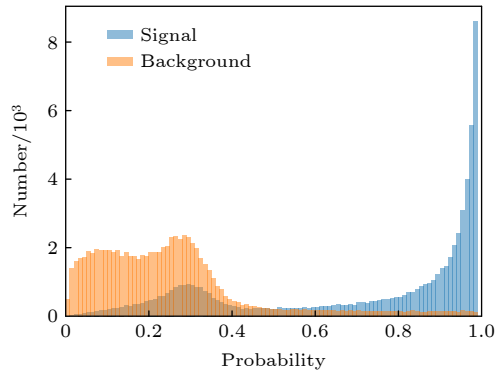


图 5 CNN 3 信号神经元对于信号(橘色)与背景(蓝色)的输出分布

Fig. 5. Distribution of the signal neuron of the CNN 3 on signal and background samples.

注图片,如图 6 和图 7 所示. 结合最优的信号与最差的背景来看,信号的中心是一个横向动量很高的亚喷注,或者具有两个亚喷注的结构. 相比之下,最优的背景显示出一种近乎随机的分布,图 7(a)具有两个亚喷注的结构,但从图中来看与信号的两个亚喷注的距离不同. 图 7(b)和图 7(c)中显示了多个亚喷注的结构并且相互各异,图 6(d)—(f)也显示出这种距离较为随机的双亚喷注结构. 总体来说,背景喷注的结构更加随机,不只是两个,多个亚喷注也可能出现,并且对比亚喷注本身,背景更加分散,信号更加集中. 除了最好与最差的喷注图片,将四种可能的识别情况: 正确识别成信号的信

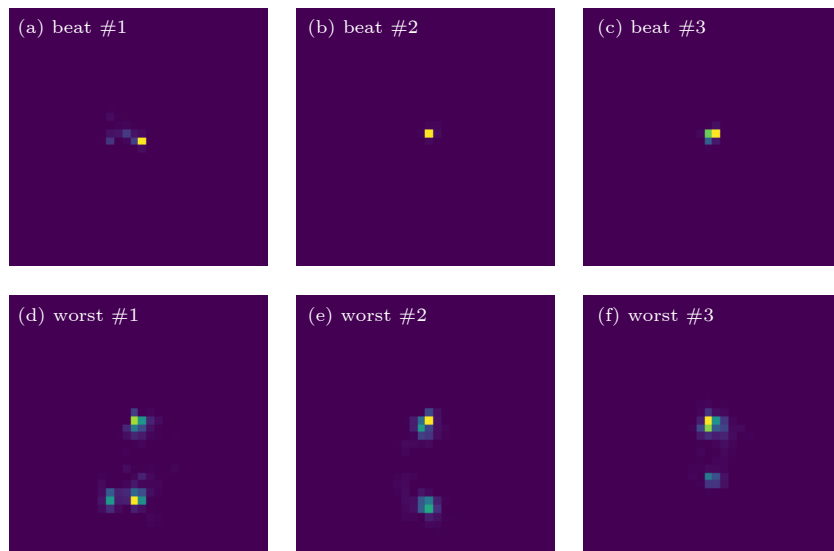


图 6 最优与最差的信号喷注图

Fig. 6. The best and the worst signal jet images.

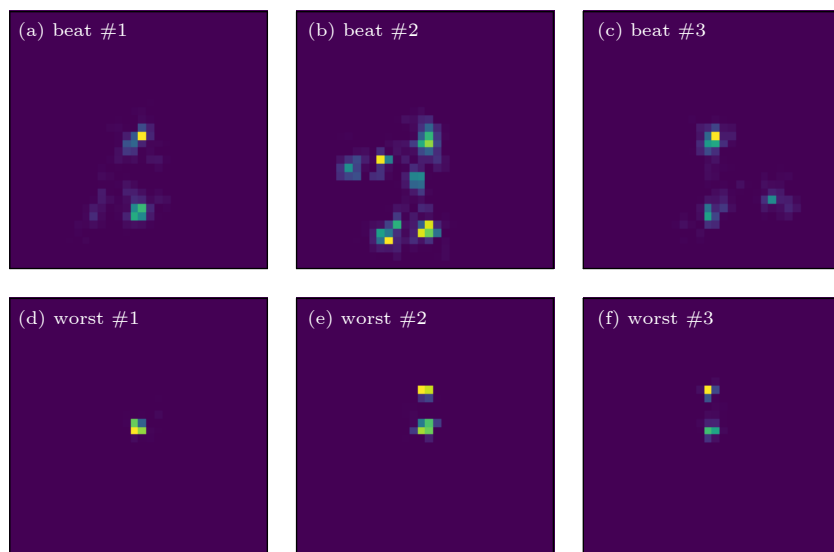


图 7 最优与最差的背景喷注图

Fig. 7. The best and the worst background jet images.

号、错误识别成背景的信号、正确识别成背景的背景和错误识别成信号的背景做成一个混淆矩阵, 来对不同类别的分辨效果进行探究. 图 8 展示了我们得到的混淆矩阵, 纵坐标代表着真实的类别, 横坐标代表着模型预测的类别. 除了大部分识别正确的喷注以外, 发现背景识别的准确率远大于信号, 信号中有将近 23% 的喷注错误地识别成了背景, 这意味着对于背景的 QCD 喷注, 模型对于来自 Z 玻色子衰变的喷注更加不确定. 由图 1 与图 7 得到这样的线索: 信号与背景的特征有很大部分是重合的, 我们训练的模型倾向于将这样重合的特征, 例如不固定  $\Delta R$ , 归为背景喷注, 而只有具有显著的单喷注或者集中的双喷注结构识别为信号, 这样的模型对于信号更加“保守”. 在实际的实验中, 信号事件的数量远远小于无关的事件数, 这就导致了信号喷注占的只是很小的一部分, 所以这样保守的模型可能会过滤掉少数的感兴趣的信号, 相反的过于激进的模型可能会将无关的背景识别成信号, 产生误导. 如何设计与训练模型才能达到在保守与激进之间的平衡, 是将来研究的重点.

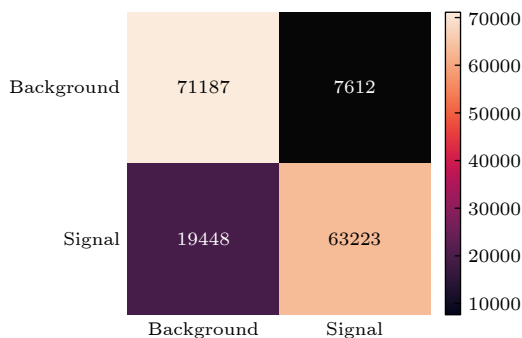


图 8 CNN 3 在测试集上的混淆矩阵, 其中纵坐标代表喷注图的真实类别, 横坐标代表模型预测的类别

Fig. 8. Confusion matrix of the CNN 3 on the test set. The true label is on the vertical axis, and the predicted label is on the horizontal axis.

## 5 结论

本文探索了把卷积神经网络应用于喷注识别任务上的潜力. 通过把喷注投影到  $\eta-\phi$  平面上, 得到了喷注的图片表示, 将它作为输入放进不同深度的卷积神经网络中. 训练测试得到的结果显示, 我们设计的 CNN 3 具有最好的识别效果, 并且在相同的表现下结构是最精简的. 与之相比较的是, 将产生的喷注的横向动量、质量、亚喷注之间的距离

作为特征输入的增强决策树, 设计的三种决策树取得了最差的分辨能力, 并且相互之间没有差异, 可能的原因是输入的种类不多. 通过找到最优与最差的信号与背景喷注, 总结了信号与背景的特点. 利用每个类别的识别情况, 做出了模型的混淆矩阵, 发现训练出的模型偏向于保守——更容易将信号错误地识别成为背景, 虽然它识别的结果相比于激进的模型来说更加可靠, 但是也更容易漏掉真实实验中数量很少的信号事件. 本文为之后设计模型以达到保守与激进的平衡, 贡献了一个基准, 为以后的评估模型提供了一种思路.

## 参考文献

- [1] Kogler R, Nachman B, Schmidt A, Asquith L, Winkels E, Campanelli M, Delitzsch C, Harris P, Hinzmann A, Kar D, McLean C, Pilot J, Takahashi Y, Tran N, Vernieri C, Vos M 2019 *Rev. Mod. Phys.* **91** 045003
- [2] Kasieczka G, Plehn T, Butter A, Crammer K, Debnath D, Dillon B M, Fairbairn M, Faroughy D A, Fedorko W, Gay C, Gouskos L, Kamenik J F, Komiske P, Leiss S, Lister A, Macaluso S, Metodiev E, Moore L, Nachman B, Nordström K, Parkes J, Qu H, Rath Y, Rieger M, Shih D, Thompson J, Varma S 2019 *SciPost Phys.* **7** 014
- [3] Larkoski A J, Moutl I, Nachman B 2020 *Phys. Rep.* **841** 1
- [4] Lai Z, Guo L, Li X Z, Dang W J 2013 *Acta Phys. Sin.* **62** 184207 (in Chinese) [来志, 郭亮, 李小珍, 党文佳 2013 *物理学报* **62** 184207]
- [5] Yang Z X, Gao Z R, Sun X F, Cai H L, Zhang F M, Wu X S 2019 *Acta Phys. Sin.* **68** 210502 (in Chinese) [杨自欣, 高章然, 孙晓帆, 蔡宏灵, 张凤鸣, 吴小山 2019 *物理学报* **68** 210502]
- [6] Xu Q W, Wang P P, Zeng Z J, Huang Z B, Zhou X X, Liu J M, Li Y, Chen S Q, Fan D Y 2020 *Acta Phys. Sin.* **69** 014209 (in Chinese) [徐启伟, 王佩佩, 曾镇佳, 黄泽斌, 周新星, 刘俊敏, 李瑛, 陈书青, 范滇元 2020 *物理学报* **69** 014209]
- [7] Cogan J, Kagan M, Strauss E, Schwartzman A 2015 *J. High Energy Phys.* **2015** 118
- [8] Almeida L G, Backović M, Cliche M, Lee S J, Perelstein M 2015 *J. High Energy Phys.* **2015** 86
- [9] Chen Y C J, Chiang C W, Cottin G, Shih D 2020 *Phys. Rev. D* **101** 053001
- [10] Qu H, Gouskos L 2020 *Phys. Rev. D* **101** 056019
- [11] Diefenbacher S, Frost H, Kasieczka G, Plehn T, Thompson J 2020 *SciPost Phys.* **8** 023
- [12] Fraser K, Schwartz M D 2018 *J. High Energy Phys.* **2018** 93
- [13] Macaluso S, Shih D 2018 *J. High Energy Phys.* **2018** 121
- [14] Lin J, Freytsis M, Moutl I, Nachman B 2018 *J. High Energy Phys.* **2018** 101
- [15] Komiske P T, Metodiev E M, Schwartz M D 2017 *J. High Energy Phys.* **2017** 110
- [16] Kasieczka G, Plehn T, Russell M, Schell T 2017 *J. High Energy Phys.* **2017** 6
- [17] Baldi P, Bauer K, Eng C, Sadowski P, Whiteson D 2016 *Phys. Rev. D* **93** 094034
- [18] Oliveira L, Kagan M, Mackey L, Nachman B, Schwartzman A 2016 *J. High Energy Phys.* **2016** 69
- [19] Guest D, Collado J, Baldi P, Hsu S C, Urban G, Whiteson D

- 2016 *Phys. Rev. D* **94** 112002
- [20] Butter A, Kasieczka G, Plehn T, Russell M 2018 *SciPost Phys.* **5** 028
- [21] Kasieczka G, Kiefer N, Plehn T, Thompson J 2019 *SciPost Phys.* **6** 069
- [22] Erdmann M, Geiser E, Rath Y, Rieger M 2019 *J. Instrum.* **14** P06006
- [23] Abdughani M, Ren J, Wu L, Yang J M 2019 *J. High Energy Phys.* **2019** 55
- [24] Komiske P T, Metodiev E M, Thaler J 2019 *J. High Energy Phys.* **2019** 121
- [25] Sjöstrand T, Ask S, Christiansen J R, Corke R, Desai N, Ilten P, Mrenna S, Prestel S, Rasmussen C O, Skands P Z 2015 *Comput. Phys. Commun.* **191** 159
- [26] Cacciari M, Salam G P, Soyez G 2012 *Eur. Phys. J. C* **72** 1896
- [27] Cacciari M, Salam G P, Soyez G 2008 *J. High Energy Phys.* **2008** 63
- [28] Krohn D, Thaler J, Wang L T 2010 *J. High Energy Phys.* **2010** 84

## Tag Z boson jets via convolutional neural networks\*

Li Jing Sun Hao<sup>†</sup>

(School of Physics, Dalian University of Technology, Dalian 116024, China)

( Received 18 September 2020; revised manuscript received 13 November 2020 )

### Abstract

The jet tagging task in high-energy physics is to distinguish signals of interest from the background, which is of great importance for the discovery of new particles, or new processes, at the large hadron collider. The energy deposition generated in the calorimeter can be seen as a kind of picture. Based on this notion, tagging jets initiated by different processes becomes a classic image classification task in the computer vision field. We use jet images as the input built on high dimensional low-level information, energy-momentum four-vectors, to explore the potential of convolutional neural networks (CNNs). Four models of different depths are designed to make the best underlying useful features of jet images. Traditional multivariable method, boosted decision tree (BDT), is used as a baseline to determine the performance of networks. We introduce four observable quantities into BDTs: the mass, transverse momenta of fat jets, the distance between the leading and subleading jets, and N-subjettiness. Different tree numbers are adopted to build three kinds of BDTs, which is intended to have variable classifying abilities. After training and testing, the results show that the CNN 3 is the neatest and most efficient network under the design of stacking convolutional layers. Deepening the model could improve the performance to a certain extent but it is unable to work all the time. The performances of all BDTs are almost the same, which is possibly due to a small number of input observable types. The performance metrics show that the CNNs outperform the BDTs: the background rejection efficiency increases up to 150% at 50% signal efficiency. Besides, after inspecting the best and the worst samples, we conclude the characteristics of jets initiated by different processes: jets obtained by Z boson decays tend to concentrate in the center of jet images or have a clear differentiable substructure; the substructures of jets from general quantum chromodynamics processes have more random forms and not only just have two jets. As the final step, the confusion matrix of the CNN 3 indicate that it comes to be kind of conservative. Exploring the way of keeping the balance between conservative and radical is our goal in the future work.

**Keywords:** decays of Z bosons, quarks, gluons, neural network

**PACS:** 13.38.Dg, 14.65.-q, 14.70.Dj, 87.85.dq

**DOI:** 10.7498/aps.70.20201557

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 11675033, 12075043).

<sup>†</sup> Corresponding author. E-mail: [haosun@dlut.edu.cn](mailto:haosun@dlut.edu.cn)